Quality Evaluation of Static Point Clouds encoded using MPEG Codecs

Stuart Perry, Huy Phi Cong
Faculty of Engineering and IT
University of Technology Sydney
Sydney, Australia
Stuart.Perry@uts.edu.au
CongHuy.Phi@student.uts.edu.au

Luís A. da Silva Cruz
Dep. Elect. and Comp. Eng., Univ. of Coimbra
Instituto de Telecomunicações
Portugal
lcruz@deec.uc.pt

João Prazeres
Instituto de Telecomunicações
Portugal

Manuela Pereira, Antonio Pinheiro
Universidade da Beira Interior
Instituto de Telecomunicações
Portugal
mpereira@di.ubi.pt
pinheiro@ubi.pt

Emil Dumic
Department of Electrical Engineering
University North
Croatia
emil.dumic@unin.hr

Evangelos Alexiou,Touradj Ebrahimi
Multimedia Signal Processing Group
École Polytechnique Fédérale de Lausanne
Lausanne, Switzerland
evangelos.alexiou@epfl.ch
touradj.ebrahimi@epfl.ch

***Abstract*—This paper presents a quality evaluation of the point cloud codecs recently standardised by the MPEG committee. A subjective experiment was designed to evaluate these codecs performance in terms of bit rate versus perceived quality. Four laboratories with experience with such studies carried out the subjective evaluation. Although the exact setups of the different laboratories were not the same, the obtained MOS results exhibit a high correlation between them, confirming reliability and repeatability of the proposed assessment protocol. The study also confirmed MPEG V-PCC as a superior compression solution for static point clouds when compared to MPEG G-PCC. Finally, a benchmark of the most popular point cloud metrics was performed based on the subjective results. The point2plane metric using the mean square error as a distance measure was revealed to have the best correlation with subjective scores, closely followed by the point2point, also using the mean square error. As both metrics produce high correlation results, it can be concluded that they can be used for quality assessment of MPEG codecs.**

***Index Terms*—Point Clouds, Subjective Quality Evaluation, Coding**

I. Introduction

Point clouds have recently attracted a strong interest in volumetric representation of visual information. Usually, point clouds lead to huge amounts of data, making its efficient compression a very important issue. However, lossy compression requires reliable quality models in order to provide bitrate reduction without excessively compromising the visual quality.

There are several studies in the literature aiming at the definition of frameworks and protocols for subjective evaluation of point clouds. Some of them are focused on geometry-only contents [1], [2], while others are assessing prior encoding solutions with inferior performance with respect to the current state-of-the-art [3]–[5], or degradations introduced by a single codec [6], [7] leading to limited generalizability.

This paper reports a study on subjective and objective quality evaluation of the state-of-the-art MPEG codecs for point clouds, namely, Video-based Point Cloud Compression (V-PCC) and Geometry-based Point Cloud Compression (G-PCC).

In the following, the subjective assessment methodology used in this study will be described. Then the subjective results obtained based on the above methodology are described and analysed. Finally, a benchmark of popular objective metrics for point clouds is reported, followed by a detailed discussion of the conclusions of this study.

II. Subjective Quality Assessment

A. Content

In this study, a dataset of 6 static point clouds was used with both texture and geometry information, namely, *Long Dress*, *Loot*, *Soldier*, *Red and Black*, which can be found in [8], [9] and *Ricardo10* and *Sarah9*, available at [8]. The models represent human figures with up to about 1 million points. Frontal views of the reference point clouds are illustrated in Fig. 1. These point clouds were encoded using MPEG G-PCC and V-PCC coders following the procedures described in the next sections.

B. Point Cloud Compression with MPEG G-PCC

In MPEG G-PCC [10], there are two encoding modules integrated to compress geometry information, namely, Octree and Triangle Soup (TriSoup) [11], [12]. The first approach is based on an Octree decomposition, which is regulated through the positionQuantizationScale parameter (herein referred as *Depth*). The second approach is based on a surface reconstruction using triangular primitives, after enclosing the model in an octree structure. The octree can be adjusted through the *Depth*

(a) *Long Dress* (b) *Loot* (c) *Soldier* (d) *Red and Black*

(e) *Ricardo10* (f) *Sarah9*

Fig. 1. Frontal views of each point cloud.

parameter, while the size of the block on which the triangular surface approximation is applied, is determined through the trisoup_node_size_log2 parameter (herein referred as *Level*). Octree-based encoding leads to regular down-sampling of the input points, whereas the TriSoup-based encoding leads to the presence of holes in the form of triangles. The latter approach is advised to be used only with high-density point clouds.

The geometry of a model is initially encoded, using any of the above solutions, and then decoded in order to define the shape over which the color will lie. In particular, the color attributes are associated to the output points (after geometric compression) through a re-colouring step that takes place and uses the color values of the original model. The color information can then be encoded using two different approaches, namely, RAHT [13], and Prediction-plus-Lifting (i.e., Lifting). The first alternative is based on the 3D Haar transform, whereas the second is based on prediction of a color value from its neighbors. In [14], it has been shown that by using the MPEG point cloud coding Common Test Conditions (CTC) [15], the observers tend to prefer the Lifting codec over the RAHT counterpart. Thus, in this study we only consider the Lifting codec, in order to reduce the parameter space of our experiment, while retaining the most diverse types of artifacts.

Five out of six encoding configurations specified in the CTC (R01-R06) were used to encode the contents using the G-PCC version 7 [16]. Specifically, we used R01, R02, R04, R05 and R06 to define 5 quality levels (Q01-Q05) that span from very low to very high. Every content was encoded using the exact settings specified in the CTC, except for *Sarah9* with the TriSoup module. In the latter case, a *Depth* value of 0.5 was used instead of 1, which was applied on the rest of the contents. This difference was made to account for the higher sparsity of this particular content. Yet, the same *Level* values were employed for the corresponding quality levels. Moreover, it should be noted that the *seq_lod* and the *seq_dist2* parameters of the Lifting were set as 12 and 3, respectively, for every content.

C. Point Cloud Compression with MPEG V-PCC

MPEG V-PCC [11], [17] encodes point clouds using an approach based on projecting each point cloud onto a set of planes (usually six) followed by 2D encoding of the projections. The projections are represented by three sets of information; texture patches, depth information and an occupancy map. The texture patches are packed to create a 2D image with as few pixels as possible which is then encoded using legacy video encoding methods. The distances from the projection plane to the points that project onto each texture patch pixel are represented by depth maps, also structured as patches matching the texture patches. The depth information is encoded using 2D video encoding methods too. The occupancy map is a 2D binary field that indicates which pixels in the 2D patch composite images (texture and depth) contain meaningful information and is encoded using a form of spatial quantization together with raster scanning and entropy encoding. The current MPEG V-PCC test model, TMC2 [18], uses HEVC to encode the sequence of projection images but other video encoders can be used.

To prepare the V-PCC encoded test contents, version 8 of the reference software was used with encoding parameters as defined in the MPEG CTC document [15]. The encoding condition selected was **C2**, *Lossy Geometry – Lossy Attributes*, and since we are encoding static point clouds, the coding mode was All Intra (AI). Six rate points were selected, five borrowed from the MPEG CTC (R01-R05 from low to high quality) and an additional one defined on purpose for this study to yield a lower quality than R01 and comparable to the lowest quality rate point used in the G-PCC.

D. Evaluation Methodology

The subjective experiments were conducted in 4 different laboratories: University of Beira Interior (UBI), Covilhã, Portugal, University of Coimbra (UC), Coimbra, Portugal, University North (UNIN), Varaždin, Croatia and University of Technology Sydney (UTS), Australia. The conditions of every test environment were adjusted to follow the ITU-R Recommendation BT.500-13 [19]. The equipment and viewing conditions used in each laboratory are described in Table I. A passive subjective evaluation methodology was applied using a customized version of the MPV video player [20]. In particular, the evaluated point clouds were shown to the subjects through video sequences of 60 fps with a total duration of 12 seconds, and the participants were able to provide their scores after the completion of the playback animation. In the video sequences, the point clouds were rotated around a vertical axis passing through the center of each model, allowing observation from different viewing angles.

The simultaneous Double Stimulus Impairment Scale (DSIS) test method was adopted with a 5-level rating scale

(1 - *very annoying*, 2 - *annoying*, 3 - *slightly annoying*, 4 - *perceptible, but not annoying*, 5 - *imperceptible*), including a hidden reference for sanity check. Both the reference and the degraded stimuli were simultaneously shown to the observer side-by-side, and every subject rated the visual quality of the processed with respect to the reference stimulus, which was clearly identified. To avoid biases, in half of the individual evaluations, the reference was placed on the right and the degraded content on the left side of the screen, and vice-versa for the rest. Also, particular care was taken to not to present the same content consecutively.

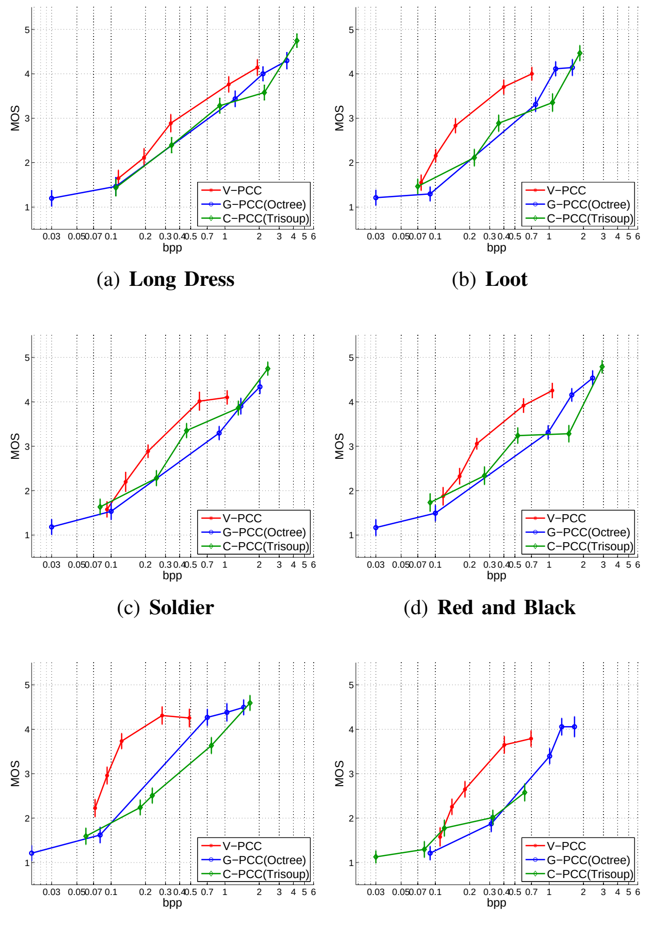
At the beginning of each individual evaluation, a training session took place using a point cloud not included in the evaluation test set, in order to familiarize the subjects with the artifacts under assessment. The content used for training was not one of the contents used for the experiment as shown in Fig. 1, but rather another example of content similar to the *Sarah9* content. The training was performed using 8 animated video sequences. For the training contents, 3 different levels of degradation were presented together with the original video sequences, in order to illustrate the range of visible distortions.

A total of 97 scores were obtained per evaluation session, considering that each subject assessed 6 test models degraded by 3 compression schemes at 5 quality levels, plus hidden references. Outlier detection based on ITU-R Recommendation BT.500-13 [19] was applied to the collected scores from every test laboratory, separately. In every case, no outliers were found. Thus, the mean opinion scores (MOS) and the 95% Confidence Intervals (CIs), assuming a Student's t-distribution were computed on every set of scores. In Table II, we report observers demographical information, per test laboratory.

III. Results

A. Subjective Scores

In Figure 2, the MOS against bitrate is depicted per content, after aggregating the scores collected from each participating laboratory. The bitrate is measured in bpp and is computed as the total number of bits required for a particular stimulus divided by the number of input points from the corresponding content. The scores were pooled together given that the MOS obtained from the laboratories were strongly correlated, as will be confirmed in the next section.



(a) **Long Dress** (b) **Loot** (c) **Soldier** (d) **Red and Black** (e) **Ricardo10** (f) **Sarah9**

Fig. 2. Subjective scores against quality levels for all laboratories and per compression method.

B. Comparison between Subjective Scores from different Labs

To confirm the validity of the subjective scores that were obtained from the participating test laboratories, the Pearson Correlation Coefficient (PCC), the Spearman Rank Order Correlation Coefficient (SROCC), the Root-Mean Squared Error (RMSE) and the Outlier Ratio (OR) are computed, to measure the linearity, monotonicity, accuracy and consistency of the results, respectively. All results are presented on Table III and for two cases, without applying any fitting function and applying linear fitting. A high correlation between results obtained from different laboratories can be observed, as the PCC and SROCC values are always higher than 0.97.

Example scatter plots with the MOS obtained in different laboratories against others are presented in Figure 3, for linear fitting function. The correlation is high, confirming the results of Table III.

TABLE I
EQUIPMENT INFORMATION AND VIEWING DISTANCE PER LABORATORY.

	Monitor	Inches	Resolution	View Distance
UBI	Eizo ColorEdge CG318-4K	31.1"	4096x2160	1.2m (±15 cm)
UC	Monitor: Sony KD-49X8005C	49"	3840x2160	1.8 m (FV ±30 cm)
UNIN	Sony TV KD-55x8505C	55"	3840x2160	1.5 m (FV ±15 cm)
UTS	Eizo ColorEdge CG318-4K	31.1"	4096x2160	1.2 m (±15 cm)

TABLE II
SUBJECTS INFORMATION PER LABORATORY.

	Males	Females	Overall	Age span	Average age
UBI	7	9	16	19-32	22
UC	7	8	15	18-54	28
UNIN	10	5	15	19-59	29
UTS	21	6	27	21-47	32

TABLE III
CONSISTENCY ACROSS LABS (BOLD TEXT REPRESENTS THE ASSUMED GROUND TRUTH).

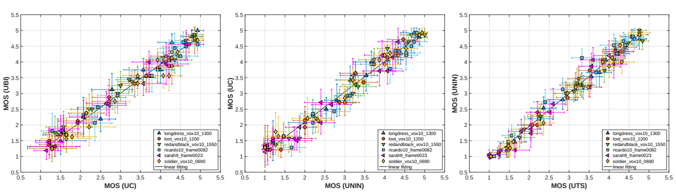
	Fitting	PCC	SROCC	RMSE	OR
UBI vs UC	no	0.987	0.980	0.213	0.041
	linear	0.987	0.980	0.183	0.052
UBI vs UNIN	no	0.984	0.982	0.254	0.093
	linear	0.984	0.982	0.201	0.052
UBI vs UTS	no	0.984	0.978	0.225	0.093
	linear	0.984	0.978	0.206	0.072
UC vs UBI	no	0.987	0.980	0.213	0.104
	linear	0.987	0.980	0.198	0.093
UC vs UNIN	no	0.986	0.982	0.219	0.072
	linear	0.986	0.982	0.206	0.072
UC vs UTS	no	0.989	0.987	0.200	0.145
	linear	0.989	0.987	0.183	0.072
UNIN vs UBI	no	0.984	0.982	0.254	0.322
	linear	0.984	0.982	0.222	0.260
UNIN vs UC	no	0.986	0.982	0.219	0.187
	linear	0.986	0.982	0.212	0.177
UNIN vs UTS	no	0.990	0.987	0.193	0.166
	linear	0.990	0.987	0.177	0.156
UTS vs UBI	no	0.984	0.978	0.225	0.270
	linear	0.984	0.978	0.212	0.208
UTS vs UC	no	0.989	0.984	0.218	0.261
	linear	0.989	0.984	0.183	0.196
UTS vs UNIN	no	0.988	0.986	0.215	0.177
	linear	0.988	0.986	0.184	0.140

C. Benchmark of objective metrics.

The point-to-point (po2point) and point-to-plane (po2plane) metrics were used to estimate geometric distortions [21], using the Mean Squared Error (MSE) and the Hausdorff distance as the geometric error measure. The geometry PSNR ratio is also computed as the max distance of nearest neighbors divided by the squared geometric error value, as defined in [21], for the two considered distances, MSE and Hausdorff.

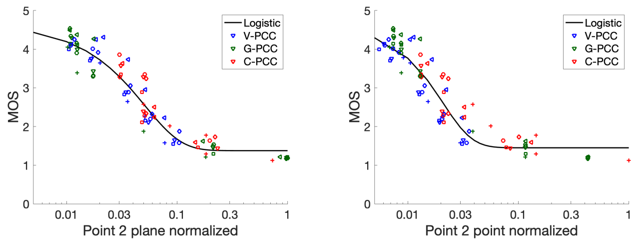
The MSE on the YCbCr representation was also employed to estimate the color degradations, after converting the default RGB to the YCbCr colorspace, following the ITU-R Recommendation BT.709-3 [22]. Moreover, the plane-to-plane (pl2plane) metric [23] is employed using a simple average, and the MSE distance. For each stimulus, the normal vectors were estimated using a quadric fitting function on neighborhoods of fixed radius (i.e., 5), as implemented in CloudCompare [24] software. Finally, the symmetric error was used to obtain a total distortion value from each metric; that is, the maximum error after setting both the original and the distorted point cloud as a reference.

To compare the objective scores against the subjective ground truth, the performance indexes proposed in the Recommendation ITU-T P.1401 [25] are employed. Specifically, the PCC, SROCC, RMSE and OR were issued on pairs of MOS and predicted MOS, to measure the performance of each metric. The predicted MOS for every objective metric, was obtained after applying the logistic fitting function on the objective scores. Based on our results as presented in Table IV, the best-performing metrics found to be the po2plane_MSE, closely followed by the metric po2point_MSE. The Scatter plots with the MOS against the two most performing metrics, po2plane_MSE and po2point_MSE are presented in Fig. 4, for a logistic fitting function.



(a) **UBI** vs UC (b) **UC** vs UNIN (c) **UNIN** vs UTS

Fig. 3. Linear fitting, for correlation evaluation between the four participating laboratories results (Bold text represents the assumed ground truth).



(a) po2plane_MSE vs MOS (b) po2point_MSE vs MOS

Fig. 4. Logistic fitting, for correlation evaluation between the best performing metrics against MOS. Each symbol represents the results of each content (▽,□,○,◇, ◁, + in the order of Fig. 1).

TABLE IV
PERFORMANCE INDEXES OF OBJECTIVE QUALITY METRICS USING SYMMETRIC ERROR AND LOGISTIC FITTING.

Metric	PCC	SROCC	RMSE	OR
po2point_MSE	0.946	0.934	0.368	0.666
po2plane_MSE	**0.959**	**0.951**	**0.321**	**0.577**
PSNR_po2point_MSE	0.868	0.855	0.540	0.752
PSNR_po2plane_MSE	0.913	0.910	0.443	0.588
po2point_HAU	0.401	0.531	1.045	0.844
po2plane_HAU	0.534	0.613	0.966	0.877
PSNR_po2point_HAU	0.548	0.456	0.911	0.870
PSNR_po2plane_HAU	0.580	0.547	0.887	0.847
color_Y_MSE	0.876	0.892	0.551	0.766
color_Cb_MSE	0.683	0.694	0.834	0.844
color_Cr_MSE	0.594	0.616	0.918	0.844
color_Y_PSNR	0.887	0.892	0.525	0.688
color_Cb_PSNR	0.693	0.694	0.822	0.844
color_Cr_PSNR	0.626	0.617	0.890	0.855
pl2plane_AVG	0.922	0.910	0.439	0.600
pl2plane_MSE	0.925	0.912	0.432	0.611

IV. CONCLUSIONS

A quality evaluation of the MPEG point cloud codecs is reported, unveiling the very high compression performance of MPEG V-PCC when used for static point clouds coding. The methodology followed revealed to be reliable and repeatable by producing consistent results in four different laboratories. Moreover, several point cloud quality metrics have a very high correlation with the MOS obtained in the subjective evaluation. In particular, the po2plane metric with distance computed using the MSE results in the highest correlations with PCC of 0.959 and SROCC of 0.951. The subjective data and test materials are available online[1].

[1] http://emergimg.di.ubi.pt/icip2020PC

References

[1] E. Alexiou and T. Ebrahimi, "On the performance of metrics to predict quality in point cloud representations," in *Proceedings of SPIE*, ser. Applications of Digital Image Processing XL, vol. 10396, Aug. 2017.

[2] E. Alexiou, T. Ebrahimi, M. V. Bernardo, M. Pereira, A. Pinheiro, L. A. D. S. Cruz, C. Duarte, L. G. Dmitrovic, E. Dumic, D. Matkovic, and A. Skodras, "Point cloud subjective evaluation methodology based on 2d rendering," in *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*, May 2018, pp. 1–6.

[3] A. Javaheri, C. Brites, F. Pereira, and J. Ascenso, "Subjective and objective quality evaluation of compressed point clouds," in *2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP)*, Oct. 2017, pp. 1–6.

[4] L. A. da Silva Cruz, E. Dumić, E. Alexiou, J. Prazeres, R. Duarte, M. Pereira, A. Pinheiro, and T. Ebrahimi, "Point cloud quality evaluation: Towards a definition for test conditions," in *2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*, June 2019, pp. 1–6.

[5] S. Perry, A. Pinheiro, E. Dumić, and L. A. da Silva Cruz, "Study of subjective and objective quality evaluation of 3d point cloud data by the jpeg committee," in *Electronic Imaging, Image Quality and System Performance XVI*, vol. 7, January 2019, pp. 312–1–312–7.

[6] E. M. Torlig, E. Alexiou, T. A. Fonseca, R. L. de Queiroz, and T. Ebrahimi, "A novel methodology for quality assessment of voxelized point clouds," in *Proceedings of SPIE*, ser. Applications of Digital Image Processing XLI, vol. 107520I, Sept. 2018.

[7] E. Zerman, P. Gao, C. Ozcinar, and A. Smolić, "Subjective and objective quality assessment for volumetric video compression," 2018.

[8] JPEG Pleno Database, https://jpeg.org/plenodb/. [Online]. Available: https://jpeg.org/plenodb/

[9] E. d'Eon, B. Harrison, T. Myers, and P. A. Chou, "8i voxelized full bodies – a voxelized point cloud dataset," ISO/IEC JTC1/SC29/WG1 input document M74006 and ISO/IEC JTC1/SC29/WG11 input document m40059, Geneva, Switzerland, Tech. Rep., January 2017.

[10] MPEG 3DG, "G-PCC Codec Description v5," ISO/IEC JTC1/SC29/WG11 N18891, Geneva, CH, October 2019.

[11] S. Schwarz, M. Preda, V. Baroncini, M. Budagavi, P. Cesar, P. A. Chou, R. A. Cohen, M. Krivokuća, S. Lasserre, Z. Li, J. Llach, K. Mammou, R. Mekuria, O. Nakagami, E. Siahaan, A. Tabatabai, A. M. Tourapis, and V. Zakharchenko, "Emerging mpeg standards for point cloud compression," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 9, no. 1, pp. 133–148, March 2019.

[12] E. Pavez, P. A. Chou, R. L. de Queiroz, and A. Ortega, "Dynamic polygon clouds: representation and compression for vr/ar," *APSIPA Transactions on Signal and Information Processing*, vol. 7, p. e15, 2018.

[13] R. L. de Queiroz and P. A. Chou, "Compression of 3d point clouds using a region-adaptive hierarchical transform," *IEEE Transactions on Image Processing*, vol. 25, no. 8, pp. 3947–3956, Aug 2016.

[14] E. Alexiou, I. Viola, T. M. Borges, T. A. Fonseca, R. L. de Queiroz, and T. Ebrahimi, "A comprehensive study of the rate-distortion performance in mpeg point cloud compression," *APSIPA Transactions on Signal and Information Processing*, vol. 8, p. e27, 2019.

[15] MPEG 3DG, "Common test conditions for point cloud compression," ISO/IEC JTC1/SC29/WG11 N17995, Macau, CN, October 2018.

[16] ——, "G-PCC Test Model v7," ISO/IEC JTC1/SC29/WG11 N18664, Gothenburg, SE, July 2019.

[17] ——, "V-PCC Codec Description," ISO/IEC JTC1/SC29/WG11 N18892, Geneva, CH, November 2019.

[18] ——, "V-PCC Test Model v8," ISO/IEC JTC1/SC29/WG11 W18884, Geneva, CH, October 2019.

[19] ITU-R BT.500-13, "Methodology for the subjective assessment of the quality of television pictures," International Telecommunications Union, Jan. 2012.

[20] MPV video player, https://mpv.io. [Online]. Available: https://mpv.io

[21] D. Tian, H. Ochimizu, C. Feng, R. Cohen, and A. Vetro, "Evaluation metrics for point cloud compression," ISO/IEC JTC m74008, , Geneva, Switzerland, Tech. Rep., January 2017.

[22] ITU-R BT.709-3, "Bt.709 : Parameter values for the hdtv standards for production and international programme exchange," International Telecommunications Union, Feb. 1998.

[23] E. Alexiou and T. Ebrahimi, "Point Cloud Quality Assessment Metric Based on Angular Similarity," in *2018 IEEE International Conference on Multimedia and Expo (ICME)*, July 2018, pp. 1–6.

[24] 3D point cloud and mesh processing software Open Source Project. CloudCompare. [Online]. Available: http://www.danielgm.net/cc/

[25] ITU-T P.1401, "Methods, metrics and procedures for statistical evaluation, qualification and comparison of objective quality prediction models," International Telecommunication Union, Jul. 2012.